Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

NeighSpec: Neighbour-Aware Spectral Alignment for Cross-Subject EEG and ECG Classification

Zhangtao Jia
Nankai University
Haoran Li
wanch_lhr@nankai.edu.cn

Nankai University
Guikang Du
Nankai University
Jiayi Shi
Nankai University
Xuqiang Wang
State Grid Tianjin Electric Power Company (China)
Xiaoli Gong
Nankai University
Jin Zhang
Nankai University
Jordi Solé-Casals
University of Cambridge

Research Article

Keywords:

Posted Date: November 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11692825/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

NeighSpec: Neighbour-Aware Spectral Alignment for Cross-Subject EEG and ECG Classification

Zhangtao Jia[1], Haoran Li[2,3*], Guikang Du[3,4], Jiayi Shi[3,4], Xuqiang Wang[5], Xiaoli Gong[2,3], Jin Zhang[2,3], Jordi Solé-Casals[6,7,8]

[1]College of Elite Engineers, Nankai University, Tianjin, China.
[2]College of Cyber Science, Nankai University, Tianjin, China.
[3]Key Laboratory of Interventional Brain-Computer Interface and Intelligent Rehabilitation; Key Lab of Data and Intelligent System Security; Frontiers Science Center for New Organic Matter, Nankai University, Tianjin, China.
[4]College of Computer Science, Nankai University, Tianjin, China.
[5]State Grid Tianjin Electric Power Company, Tianjin, 300010, China.
[6]Data and Signal Processing Group, University of Vic–Central University of Catalonia, Vic, 08500, Spain.
[7]College of Artificial Intelligence, Nankai University, Tianjin, 300350, China.
[8]Department of Psychiatry, University of Cambridge, Cambridge, CB2 3EB, United Kingdom.

*Corresponding author(s). E-mail(s): Wanch_lhr@nankai.edu.cn;

Abstract

Background: Cross-subject generalisation in electroencephalogram (EEG) and electrocardiogram (ECG) aims to learn representations that preserve task-relevant physiological information while mitigating subject-specific variability. Existing spectral alignment methods adaptively adjust frequency components to reduce such variability. Frequency-band adaptation can reduce these variations, yet the correction of a spectral coefficient also depends on the neighbouring frequencies that define its local context. **Methods:** We propose NeighSpec, a neighbourhood-aware alignment model that derives a local spectral reference from adjacent frequency coefficients and uses a bounded relative discrepancy to guide magnitude refinement. Sample-conditioned band descriptors control gain, neighbourhood

refinement, and phase rotation. Residual temporal embedding and ordered temporal pooling support classification without requiring subject identity at inference. **Results:** Across six EEG and ECG benchmarks, NeighSpec achieves an average rank of 1.25 over 36 metric means and the highest dataset-level average on five benchmarks. Compared with published BioFormer results, macro-F1 improves by 9.93 percentage points on the BCI Competition IV 2a motor-imagery dataset and by 4.62 points on the 2b dataset. **Conclusions:** Explicit neighbourhood-relative corrections offer a compact alternative for cross-subject spectral adaptation. Their benefits are pronounced in motor-imagery classification, while analyses across datasets show that improved task performance does not imply complete subject invariance.

Keywords: biomedical time series, cross-subject generalisation, EEG, ECG, spectral neighbourhoods, frequency-domain adaptation

1 Introduction

Biomedical time-series (BTS) are temporally ordered measurements of physiological signals such as electroencephalogram (EEG) and electrocardiogram (ECG) [1–3]. Across physiological and pathological conditions, these signals exhibit distinct and structured temporal–spectral patterns that arise from underlying biological mechanisms [4, 5]. This makes BTS signals a useful source of evidence for downstream learning tasks such as disease screening, treatment monitoring, and emotion recognition [6, 7]. For example, in Alzheimer's disease, disease-related brain changes can disrupt neural synchronisation and functional connectivity, leading to systematic alterations in EEG rhythms, so that recordings from patients and healthy controls tend to exhibit two distinguishable physiological patterns [8, 9]. BTS provide a key data foundation for clinical analysis, and therefore learning on BTS is an important problem.

Cross-subject generalisation is a central challenge for deploying BTS classification models, because training is performed on recordings from a subset of subjects while evaluation targets unseen individuals [10, 11]. As a result, performance often drops at test time for two reasons. First, inter-subject variability can mask class-relevant physiological patterns in raw time-series signals, making them harder to learn [12, 13]. Even within the same class, recordings from different subjects may show large pointwise discrepancies, as illustrated in the top row of Figure 1. Second, models may exploit subject-specific shortcuts, overfitting to individual idiosyncrasies rather than learning task-relevant physiological patterns [14, 15].

Suppressing subject-specific variability is key to cross-subject generalisation in BTS [16, 17]. Prior work has approached this problem from different perspectives. First, stronger backbones, especially Transformer-based models, increase representation capacity through multi-scale temporal modelling and cross-channel interactions, implicitly reducing subject-specific variability [18]. Second, adversarial methods encourage subject-invariant representations by penalising subject predictability through adversarial learning [19, 20]. More recently, BioFormer [21] introduced spectral drift as

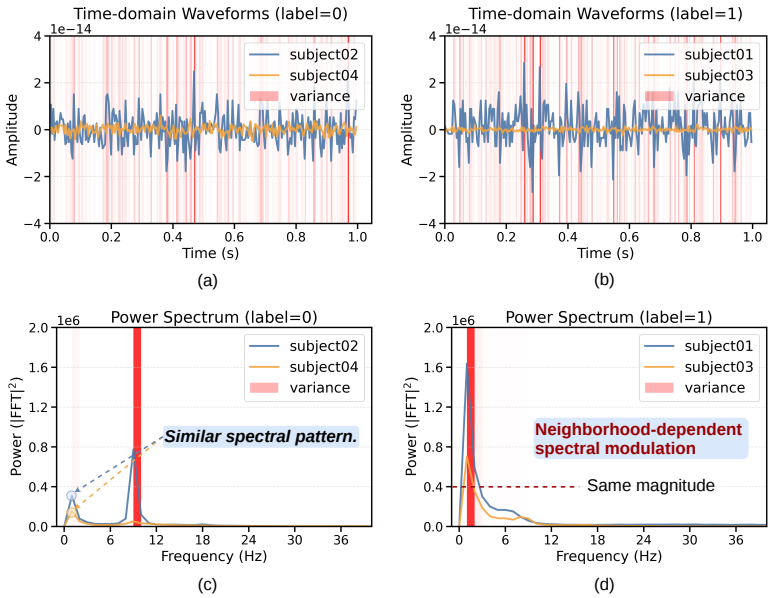

Fig. 1 Spectral variability and local frequency context in the APAVA EEG dataset. The panels contrast time-domain recordings and magnitude-spectrum examples from subjects sharing a diagnostic label. Shaded regions highlight between-subject differences; the illustration also motivates considering the neighbouring frequencies when modulating individual spectral coefficients. The amplitude and spectral-power plots have different scales and should not be interpreted as directly comparable quantities.

an explicit characterisation of subject-specific variability, using adaptive band-wise magnitude and phase modulation to align spectral representations.

However, spectral alignment raises a further question: **How should individual frequency components be modulated according to their relationships with neighbouring frequencies?** Although band-level statistics provide useful information about spectral variations, they do not fully characterise the local spectral context of each component. Frequency components with similar magnitudes may exhibit different relationships with their neighbours, and therefore require different adjustments. Without explicitly accounting for these relationships, spectral modulation may introduce unnecessary changes to task-relevant structures, potentially compromising cross-subject generalisation. Moreover, a physiological reason to examine neighbouring frequencies is that EEG oscillations occupy locally organised spectral ranges. During imagined movement, sensorimotor mu and beta rhythms can show spatially and temporally selective changes in power [22]. In Alzheimer's disease, alterations in alpha-band activity have also been associated with changes in rhythmic brain dynamics [8]. Such examples motivate modelling the local spectral context without assuming that latent frequency bins correspond directly to canonical EEG bands.

To address this issue, we investigate the role of local spectral relationships in cross-subject adaptation. Our motivation stems from the observation that physiological signals often exhibit consistent global spectral patterns while retaining local variations

across individuals. As illustrated in Figure 1, samples sharing the same label exhibit substantial time-domain differences but similar overall spectral patterns. Nevertheless, localised magnitude differences remain across subjects, as highlighted around 10 Hz in Figure 1(c). Importantly, the meaning of a frequency component depends not only on its magnitude but also on its surrounding spectral configuration. A component located at a spectral peak, along its shoulders, or within a relatively flat region may have a similar magnitude while exhibiting different local relationships. This suggests that spectral adaptation should consider the relative variation between a component and its frequency neighbourhood rather than relying solely on band-level information.

Building on this insight, we propose NeighSpec, a neighbourhood-aware spectral learning framework for cross-subject BTS classification. At its core, we introduce a Neighbourhood-Referenced Spectral Alignment (NRSA) module that explicitly incorporates local frequency relationships into spectral adaptation. Specifically, the module extracts compact band-wise statistics and integrates cross-band information to generate sample-conditioned modulation parameters. Within each band, a local spectral reference is constructed from neighbouring frequency components, and its relative discrepancy is used to guide adaptive magnitude correction. By combining band-wise gain adjustment with bounded neighbourhood-relative refinement and phase rotation, the proposed method enables controlled spectral modulation while limiting excessive local changes. We further integrate residual temporal embedding and an ordered temporal readout to capture local waveform characteristics and retain temporal information for downstream classification.

In summary, our main contributions are as follows:

(1) We highlight the importance of local spectral relationships in cross-subject adaptation and formulate spectral modulation from a neighbourhood-relative perspective, where the correction of each frequency component is guided by its surrounding spectral context.

(2) We propose NeighSpec, a neighbourhood-aware spectral learning framework that integrates neighbourhood-referenced magnitude refinement, band-wise gain adjustment, and phase rotation to achieve controlled spectral adaptation for cross-subject biomedical time-series classification.

(3) We conduct extensive evaluations on six biomedical time-series benchmarks. NeighSpec achieves the best overall average rank of 1.25 across 36 evaluation metrics and the highest average performance on five datasets, demonstrating its effectiveness for cross-subject generalisation.

2 Related Work

Temporal Representation Learning for BTS. Transformer-based architectures have become widely used in biomedical time-series analysis due to their ability to capture long-range dependencies and complex temporal patterns [18]. Recent approaches explore multi-scale representations, temporal patching, and cross-channel interactions to improve feature extraction [23, 24]. Shape-based models further investigate discriminative temporal patterns for efficient classification [25]. Although these architectures provide powerful representations, their ability to accommodate inter-subject differences

largely depends on the features learned from training subjects. In contrast, NeighSpec incorporates explicit spectral adaptation into temporal representation learning, allowing frequency-dependent variations to be adjusted before classification.

Domain Generalisation for Cross-Subject Learning. Cross-subject classification is closely related to domain generalisation, where physiological recordings from different subjects exhibit distinct data distributions. Existing approaches seek transferable representations through adversarial invariance learning, statistical alignment, and distribution perturbation [19, 26, 27]. Recent methods tailored to biomedical signals further explore subject-aware adaptation and representation alignment to mitigate individual differences [12, 28]. These approaches primarily address subject variability through representation-level objectives or adaptation strategies. Our approach focuses instead on how spectral components can be adaptively adjusted according to their frequency context, without requiring explicit subject supervision.

Frequency-Domain Modelling. Frequency-domain representations provide a natural perspective for analysing biomedical signals, where physiological information is often associated with characteristic oscillatory activity [29, 30]. Recent time-series models have incorporated spectral information to capture periodic patterns and temporal-frequency dependencies [31, 32]. Beyond extracting frequency features, spectral adaptation offers a way to account for distribution variations across individuals. However, the adjustment of a frequency component should also consider its relationship with neighbouring components, as local spectral configurations may contain distinct task-relevant information. This motivates our neighbourhood-referenced formulation, which uses local spectral discrepancies to guide bounded magnitude correction while retaining band-wise gain and phase adaptation.

BioFormer [21] introduced spectral drift as a characterisation of cross-subject variability and proposed adaptive frequency-band alignment through magnitude convolution and phase modulation. Extending this perspective, NeighSpec explicitly measures the relative discrepancy between each frequency component and its local spectral reference. The resulting formulation separates band-wise amplitude scaling from bounded neighbourhood-relative correction, enabling more controlled spectral modulation. In addition, NeighSpec adopts residual temporal embedding and an ordered temporal readout to retain temporal information throughout representation learning.

3 Problem Setup and Formulation

We study cross-subject classification for biomedical time-series data. Let $\mathcal{Q} = \{(X_i, y_i, s_i)\}_{i=1}^{N}$ denote a dataset, where $X_i \in \mathbb{R}^{T\times C}$ is a multivariate time-series segment of length T with C channels, $y_i \in \mathcal{Y}$ is the corresponding class label, and $s_i \in \mathcal{S}$ denotes the subject identity. We consider two common labelling regimes. In *subject-level* labelling, each subject is associated with a single label that is shared by all its segments. In *sample-level* labelling, each segment is labelled individually, and labels may vary within the same subject. In the cross-subject setting, subjects are partitioned into disjoint training and test sets, denoted by $\mathcal{S}_{\text{train}}$ and $\mathcal{S}_{\text{test}}$, respectively, with $\mathcal{S}_{\text{train}} \cap \mathcal{S}_{\text{test}} = \emptyset$, as shown in Figure 2. A model f_θ is trained using samples from subjects in $\mathcal{S}_{\text{train}}$ and evaluated on samples from previously unseen subjects in $\mathcal{S}_{\text{test}}$.

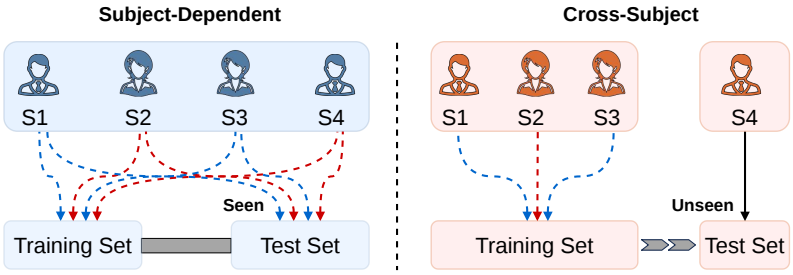

Fig. 2 Subject-dependent and cross-subject evaluation. Left: the same individual may contribute data to both training and evaluation subsets. Right: subject-disjoint splits test generalisation to individuals not seen during model fitting. Adapted from BioFormer [21].

4 Methodology

4.1 Overview

We study cross-subject classification of biomedical time series (BTS), where a model is trained on recordings from source subjects and evaluated on unseen individuals (Figure 2). Given an input segment $X \in \mathbb{R}^{T \times C}$ with T time steps and C channels, our objective is to learn representations that capture task-relevant physiological information while accommodating subject-dependent spectral variations.

To this end, we propose NeighSpec, a neighbourhood-aware spectral learning framework consisting of three stages (Figure 3). First, a residual temporal embedding extracts short-range waveform patterns without reducing temporal resolution. Next, a stack of spectral encoder blocks adapts frequency components using both band-level context and their local spectral neighbourhoods. Finally, an ordered temporal readout aggregates the encoded features while retaining coarse temporal positions for classification.

Residual temporal embedding. Physiological signals contain short-lived waveform patterns that may be weakened by early temporal aggregation. We therefore first project each observation into a D-dimensional space and use a residual temporal convolution to capture local variations while preserving the original sequence length:

$$U = XW_{\text{in}} + b_{\text{in}}, \qquad H^{(0)} = U + \text{Drop}(\text{TConv}(\text{LN}(U))) \,. \tag{1}$$

Here, TConv comprises two length-preserving one-dimensional convolutions with an intermediate GELU activation. The residual connection retains the projected signal alongside the learned local features, providing the input $H^{(0)} \in \mathbb{R}^{T \times D}$ to the spectral encoder.

4.2 Neighbourhood-Referenced Spectral Alignment

The correction of a frequency component should account for both its frequency band and its relationship with neighbouring components. Band-level information can indicate how an oscillatory region varies across samples, whereas the local neighbourhood distinguishes a coefficient near a dominant response from one in a relatively uniform

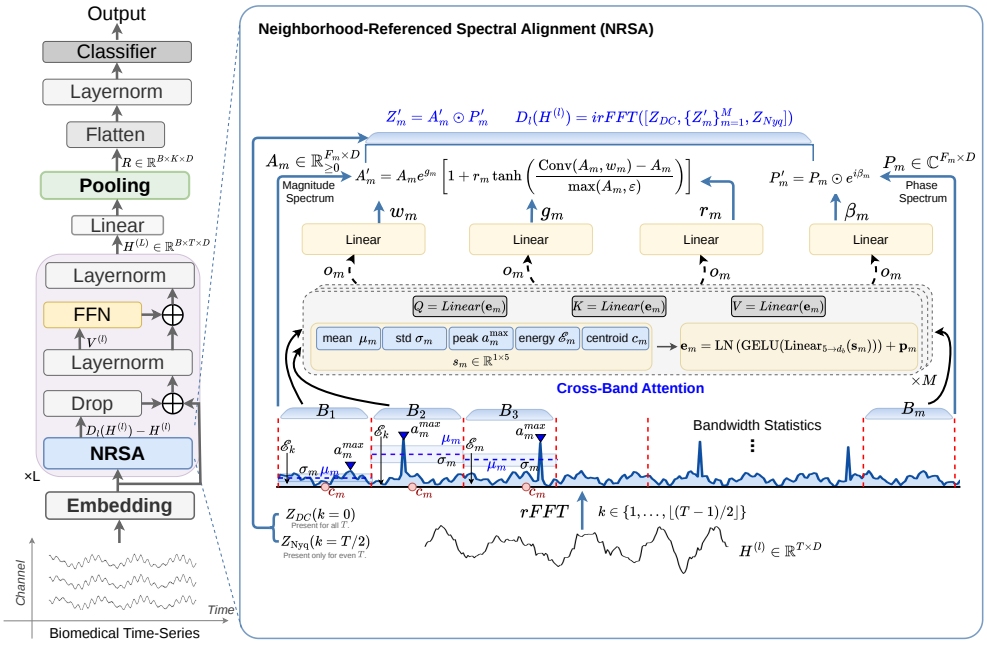

Fig. 3 Overview of NeighSpec. A residual temporal embedding extracts local waveform features, followed by neighbourhood-referenced spectral modulation within residual encoder blocks. An ordered temporal readout aggregates the resulting representations for classification.

region. We therefore construct a neighbourhood reference for each coefficient and use its relative difference to guide a bounded spectral adjustment.

Band-conditioned modulation parameters. Given a temporal representation $H \in \mathbb{R}^{T\times D}$, we obtain its complex spectrum $Z = \mathcal{F}(H)$ and magnitude $A = |Z|$ through the real Fourier transform. The non-DC frequencies, excluding the Nyquist component when present, are divided into M contiguous bands $\{B_m\}_{m=1}^{M}$. For each band, we summarise its magnitude distribution by five statistics: mean μ_m, standard deviation σ_m, peak magnitude $a_m^{\max}$, mean spectral energy $\mathcal{E}_m$, and energy-weighted centroid c_m. Specifically, the first four statistics are aggregated over the frequency bins and latent channels, while c_m indicates the relative location of spectral energy within the band. They form the descriptor $\mathbf{s}_m$, which is projected to a band embedding $\mathbf{e}_m$:

$$\begin{aligned}\mathbf{s}_m &= \big[\log(1+\mu_m),\ \log(1+\sigma_m),\ \log(1+a_m^{\max}),\ \log(1+\mathcal{E}_m),\ c_m\big],\\ \mathbf{e}_m &= \mathrm{LN}(\mathrm{GELU}(\mathrm{Linear}(\mathbf{s}_m))) + \mathbf{p}_m,\end{aligned} \tag{2}$$

where $\mathbf{p}_m$ encodes the band position. As illustrated in Figure 3, the band embeddings interact through attention, with query, key, and value projections formed from $\{\mathbf{e}_m\}_{m=1}^{M}$.

The resulting context $\mathbf{o}_m$ is mapped to four controls by separate prediction heads:

$$\begin{aligned} \mathbf{o}_m &= \mathrm{BandAttn}\big(\{\mathbf{e}_j\}_{j=1}^{M}\big)_m\,, \\ w_m &= \mathrm{Softmax}(f_w(\mathbf{o}_m))\,, \qquad g_m = G\tanh(f_g(\mathbf{o}_m))\,, \\ r_m &= \rho\tanh(f_r(\mathbf{o}_m))\,, \qquad \beta_m = \Phi\tanh(f_\beta(\mathbf{o}_m))\,. \end{aligned} \tag{3}$$

Here f_w, f_g, f_r, and f_β denote independent learnable linear heads, corresponding to the four branches in Figure 3. The three-tap kernel w_m provides a neighbourhood reference; g_m, r_m, and β_m control band-wise gain, local refinement, and phase rotation, respectively. The kernel is shared across latent channels within each band, whereas the other controls are channel-specific. The constants G, $\rho < 1$, and Φ bound their respective modulation strengths.

Neighbourhood-referenced magnitude correction. Let $A_m = |Z_m|$ denote the magnitude spectrum in band B_m. We construct a local reference $\widetilde{A}_m = \mathrm{Conv}(A_m, w_m)$ by aggregating adjacent frequency components within the same band. Instead of replacing the original magnitude with this reference, we use their relative difference to determine a controlled correction:

$$A'_m = A_m\, e^{g_m}\left[1 + r_m \tanh\left(\frac{\widetilde{A}_m - A_m}{\max(A_m, \varepsilon)}\right)\right]. \tag{4}$$

The exponential term adjusts the overall strength of each band, while the second term adapts individual coefficients according to their neighbourhoods. Importantly, these two effects can be controlled separately: when $r_m = 0$, the update reduces to $A'_m = e^{g_m} A_m$, which changes band amplitude without altering the within-band normalised magnitude profile. With $|g_m| \le G$ and $|r_m| \le \rho < 1$, the correction remains positive and its relative magnitude is bounded. Thus, neighbourhood information guides local adaptation without allowing arbitrarily large changes to individual coefficients.

Phase modulation and reconstruction. Subject-dependent variation may also affect the phase of oscillatory components. To adjust phase without coupling it to magnitude refinement, we apply a band-wise rotation to the unit-complex phase $P_m = e^{\mathrm{i}\angle Z_m}$:

$$P'_m = P_m \odot e^{\mathrm{i}\beta_m}, \qquad Z'_m = A'_m \odot P'_m. \tag{5}$$

The corrected bands are combined with the unchanged DC and Nyquist coefficients, and an inverse Fourier transform returns the adapted temporal representation $\mathcal{D}(H)$. The modulation heads for gain, local refinement, and phase are zero-initialised, allowing the spectral transform to start from the identity mapping and learn task-dependent corrections during training.

4.3 Spectral Encoder and Temporal Readout

Spectral Encoder. To progressively learn spectrally adapted representations, we integrate the NRSA module into a residual encoder followed by a feed-forward

network (FFN). Given an intermediate representation $H^{(\ell)}$, NRSA produces a spectrally corrected representation $\mathcal{D}_\ell(H^{(\ell)})$. We incorporate the spectral correction through a residual connection, followed by layer normalisation and nonlinear feature transformation:

$$\begin{aligned} V^{(\ell)} &= \mathrm{LN}\left(H^{(\ell)} + \mathrm{Drop}\left(\mathcal{D}_\ell(H^{(\ell)}) - H^{(\ell)}\right)\right), \\ H^{(\ell+1)} &= \mathrm{LN}\left(V^{(\ell)} + \mathrm{FFN}(V^{(\ell)})\right). \end{aligned} \tag{6}$$

Here, the residual branch introduces the change produced by NRSA, while the FFN captures nonlinear interactions across latent channels. Stacking L encoder layers enables progressive spectral adaptation and feature refinement.

Temporal Readout. Although spectral adaptation targets subject-dependent variations, temporal information remains important for distinguishing physiological patterns. Conventional global average pooling aggregates features across the entire sequence, potentially discarding their temporal positions. To preserve coarse temporal structure, we divide the encoded representation into K ordered temporal bins through adaptive average pooling:

$$\begin{aligned} R &= \mathrm{Pool}_K\left(\mathrm{GELU}(H^{(L)}W_t + b_t)\right), \\ \hat{\mathbf{y}} &= \mathrm{MLP}\left(\mathrm{LN}(\mathrm{vec}(R))\right). \end{aligned} \tag{7}$$

Here, $R \in \mathbb{R}^{K\times D}$ retains the relative ordering of temporal segments, and $\mathrm{vec}(\cdot)$ concatenates their representations for classification. The entire network is trained end-to-end using only the task classification objective, without requiring subject identity supervision.

5 Experiments

Datasets. We evaluate NeighSpec under a strict *cross-subject* protocol on six biomedical time-series benchmarks spanning both EEG and ECG modalities. These datasets cover two common labelling paradigms introduced in Sec. 3: (i) *subject-level labelling*, where all samples from a subject share the same class label (APAVA [33], ADFTD [34], PTB [35], PTB-XL [36]), and (ii) *sample-level labelling*, where labels are assigned independently to each sample (BCI-2a [37] and BCI-2b [38]). Table 1 summarises the dataset scale, task settings, and signal modalities.

Baselines. We compare NeighSpec with 12 attention-based and shape-based models: Transformer [39], Reformer [40], Informer [41], Autoformer [42], FEDformer [31], Non-stationary Transformer (Nonformer) [43], Crossformer [44], PatchTST [23], iTransformer [45], MTST [46], Medformer [24], and SoftShape [25]. BioFormer [21] is included as an additional baseline, with its published cross-subject results reported separately from the NeighSpec evaluations. This yields 14 methods in the complete comparison. We used the same subject-disjoint splits and random seeds as in the Medformer.

Implementation and reporting. NeighSpec results are reported as means and standard deviations over seeds 41–45. Table 2 provides Accuracy, Precision, Recall, macro-F1, AUROC, and AUPRC for all six datasets. The model is implemented

Table 1 Summary of datasets and task settings. "S-K" denotes a K-class *subject-level* classification setting, where all samples from the same subject share one label. "M-K" denotes a K-class *sample-level* classification setting, where each subject may contribute samples from multiple classes.

Dataset	Subjects	Samples	Task	Channels	Modality
APAVA	23	5,967	S-2	16	EEG
ADFTD	88	69,752	S-3	19	EEG
PTB	198	64,356	S-2	15	ECG
PTB-XL	17,596	191,400	S-5	12	ECG
BCI-2a	9	5,184	M-4	3	EEG
BCI-2b	9	6,520	M-2	3	EEG

in PyTorch and its computation is described in Section 4. Training settings, model configuration and data augmentation are given in Appendix B.3.

5.1 Cross-Subject Classification Performance

Table 2: Full cross-subject comparison with attention-based and shape-based models. Results are reported as mean ± std (%). Ranking uses the mean only. Best and second-best results are highlighted in red and blue, respectively. These include all baseline values for BCI-2a and BCI-2b and the SoftShape results reproduced by BioFormer on all six datasets.

Dataset	Model (Pub.)	Acc	Prec	Rec	F1	AUROC	AUPRC	Avg. Rank ↓
APAVA (2 classes)	Transformer [NeurIPS'17]	76.30 ± 4.72	77.64 ± 5.95	73.09 ± 5.01	73.75 ± 5.38	72.50 ± 6.60	72.23 ± 7.60	7.67
	Reformer [ICLR'20]	78.70 ± 2.00	82.50 ± 3.95	75.00 ± 1.61	75.93 ± 1.82	73.94 ± 1.40	76.04 ± 1.14	5.67
	Informer [AAAI'21]	73.11 ± 4.40	75.17 ± 6.06	69.17 ± 4.56	69.47 ± 5.06	70.46 ± 4.91	70.75 ± 5.27	10.83
	Autoformer [NeurIPS'21]	68.64 ± 1.82	68.48 ± 2.10	68.77 ± 2.27	68.06 ± 1.94	75.94 ± 3.61	74.38 ± 4.05	11.00
	FEDformer [ICML'22]	74.94 ± 2.15	74.59 ± 1.50	73.56 ± 3.55	73.51 ± 3.39	83.72 ± 1.97	82.94 ± 2.37	7.00
	Nonformer [NeurIPS'22]	71.89 ± 3.81	71.80 ± 4.58	69.44 ± 3.56	69.74 ± 3.84	70.55 ± 2.96	70.78 ± 4.08	10.83
	PatchTST [ICLR'23]	67.03 ± 1.65	78.76 ± 1.28	59.91 ± 2.02	55.97 ± 3.10	65.65 ± 0.28	67.99 ± 0.76	13.00
	Crossformer [ICLR'23]	73.77 ± 1.95	79.29 ± 4.36	68.86 ± 1.70	68.93 ± 1.85	72.39 ± 3.33	72.05 ± 3.65	9.67
	iTransformer [ICLR'24]	74.55 ± 1.66	74.77 ± 2.10	71.76 ± 1.72	72.30 ± 1.79	85.59 ± 1.55	84.39 ± 1.57	6.83
	MTST [AISTATS'24]	71.14 ± 1.59	79.30 ± 0.97	65.27 ± 2.28	64.01 ± 3.16	68.87 ± 2.34	71.06 ± 1.60	11.33
	Medformer [NeurIPS'24]	78.74 ± 0.64	81.11 ± 0.84	75.40 ± 0.66	76.31 ± 0.71	83.20 ± 0.91	83.66 ± 0.92	4.50
	SoftShape [ICML'25]	81.22 ± 0.62	85.65 ± 0.22	77.55 ± 0.87	78.71 ± 0.91	82.24 ± 2.43	83.93 ± 2.19	3.67
	BioFormer [ICML'26]	82.31 ± 1.55	87.87 ± 0.67	78.50 ± 1.94	79.77 ± 2.09	88.52 ± 2.25	88.16 ± 2.40	1.67
	NeighSpec	83.03 ± 1.36	86.35 ± 1.91	80.09 ± 2.07	81.16 ± 1.96	88.22 ± 1.82	88.49 ± 2.04	1.33
ADFTD (3 classes)	Transformer [NeurIPS'17]	50.47 ± 2.14	49.13 ± 1.83	48.01 ± 1.53	48.09 ± 1.59	67.93 ± 1.59	48.93 ± 2.02	6.17
	Reformer [ICLR'20]	50.78 ± 1.17	49.64 ± 1.49	49.89 ± 1.67	47.94 ± 0.69	69.17 ± 1.58	51.73 ± 1.94	4.83
	Informer [AAAI'21]	48.45 ± 1.96	46.54 ± 1.68	46.06 ± 1.84	45.74 ± 1.38	65.87 ± 1.27	47.60 ± 1.30	9.33
	Autoformer [NeurIPS'21]	45.25 ± 1.48	43.67 ± 1.94	42.96 ± 2.03	42.59 ± 1.85	61.02 ± 1.82	43.10 ± 2.30	13.00
	FEDformer [ICML'22]	46.30 ± 0.59	46.05 ± 0.76	44.22 ± 1.38	43.91 ± 1.37	62.62 ± 1.75	46.11 ± 1.44	11.17
	Nonformer [NeurIPS'22]	49.95 ± 1.05	47.71 ± 0.97	47.46 ± 1.50	46.96 ± 1.35	66.23 ± 1.37	47.33 ± 1.78	8.17
	PatchTST [ICLR'23]	44.37 ± 0.95	42.40 ± 1.13	42.06 ± 1.48	41.97 ± 1.37	60.08 ± 1.50	42.49 ± 1.79	14.00
	Crossformer [ICLR'23]	50.45 ± 2.31	45.57 ± 1.63	45.88 ± 1.82	45.50 ± 1.70	66.45 ± 2.03	48.33 ± 2.05	9.17
	iTransformer [ICLR'24]	52.60 ± 1.59	46.79 ± 1.27	47.28 ± 1.29	46.79 ± 1.13	67.26 ± 1.16	49.53 ± 1.21	7.00
	MTST [AISTATS'24]	45.60 ± 2.03	44.70 ± 1.33	45.05 ± 1.30	44.31 ± 1.74	62.50 ± 0.81	45.16 ± 0.85	11.67
	Medformer [NeurIPS'24]	53.27 ± 1.54	51.02 ± 1.57	50.71 ± 1.55	50.65 ± 1.51	70.93 ± 1.19	51.21 ± 1.32	3.33
	SoftShape [ICML'25]	54.60 ± 0.82	50.33 ± 1.24	49.17 ± 0.89	48.91 ± 1.13	70.08 ± 0.98	50.90 ± 1.12	4.00
	BioFormer [ICML'26]	56.73 ± 2.51	54.66 ± 2.59	54.60 ± 3.02	53.82 ± 3.03	74.74 ± 2.08	58.07 ± 2.72	1.00
	NeighSpec	54.24 ± 0.94	53.67 ± 0.60	53.49 ± 1.04	53.39 ± 0.89	72.13 ± 0.40	54.69 ± 0.71	2.17
PTB (2 classes)	Transformer [NeurIPS'17]	77.37 ± 1.02	81.84 ± 0.66	67.14 ± 1.80	68.47 ± 2.19	90.08 ± 1.76	87.22 ± 1.68	9.33
	Reformer [ICLR'20]	77.96 ± 2.13	81.72 ± 1.61	68.20 ± 3.35	69.65 ± 3.88	91.13 ± 0.74	88.42 ± 1.30	8.50
	Informer [AAAI'21]	78.69 ± 1.68	82.87 ± 1.02	69.19 ± 2.90	70.84 ± 3.47	92.09 ± 0.53	90.02 ± 0.60	6.67
	Autoformer [NeurIPS'21]	73.35 ± 2.10	72.11 ± 2.89	63.24 ± 3.17	63.69 ± 3.84	78.54 ± 3.48	74.25 ± 3.53	14.00
	FEDformer [ICML'22]	76.05 ± 2.54	77.58 ± 3.61	66.10 ± 3.55	67.14 ± 4.37	85.93 ± 4.31	82.59 ± 5.42	12.33
	Nonformer [NeurIPS'22]	78.66 ± 0.49	82.77 ± 0.86	69.12 ± 0.87	70.90 ± 1.00	89.37 ± 2.51	86.67 ± 2.38	8.33
	PatchTST [ICLR'23]	74.74 ± 1.62	76.94 ± 1.51	63.89 ± 2.71	64.36 ± 3.38	88.79 ± 1.91	83.39 ± 0.96	12.33
	Crossformer [ICLR'23]	80.17 ± 3.79	85.04 ± 1.83	71.25 ± 6.29	72.75 ± 7.19	88.55 ± 3.45	87.31 ± 3.25	7.17
	iTransformer [ICLR'24]	83.89 ± 0.71	88.25 ± 1.18	76.39 ± 1.01	79.06 ± 1.06	91.18 ± 1.16	90.93 ± 0.98	4.00
	MTST [AISTATS'24]	76.59 ± 1.90	79.88 ± 1.90	66.31 ± 2.95	67.38 ± 3.71	86.86 ± 2.75	83.75 ± 2.84	11.17
	Medformer [NeurIPS'24]	83.50 ± 2.01	85.19 ± 0.94	77.11 ± 3.39	79.18 ± 3.31	92.81 ± 1.48	90.32 ± 1.54	4.17
	SoftShape [ICML'25]	83.72 ± 0.83	86.70 ± 0.67	75.74 ± 1.28	78.42 ± 1.30	94.38 ± 0.76	93.00 ± 0.88	4.00
	BioFormer [ICML'26]	87.73 ± 1.73	88.71 ± 1.28	82.63 ± 2.80	84.71 ± 2.51	94.71 ± 0.94	93.53 ± 0.96	2.00
	NeighSpec	89.54 ± 2.05	89.89 ± 1.91	85.64 ± 3.04	87.27 ± 2.74	95.48 ± 1.29	94.67 ± 1.65	1.00

Dataset	Model (Pub.)	Acc	Prec	Rec	F1	AUROC	AUPRC	Avg. Rank ↓
PTB-XL (5 classes)	Transformer [NeurIPS'17]	70.59 ± 0.44	61.57 ± 0.65	57.62 ± 0.35	59.05 ± 0.25	88.21 ± 0.16	63.36 ± 0.29	10.75
	Reformer [ICLR'20]	71.72 ± 0.43	63.12 ± 1.02	59.20 ± 0.75	60.69 ± 0.18	88.80 ± 0.24	64.72 ± 0.47	8.17
	Informer [AAAI'21]	71.43 ± 0.32	62.64 ± 0.60	59.12 ± 0.47	60.44 ± 0.43	88.65 ± 0.19	64.76 ± 0.17	8.83
	Autoformer [NeurIPS'21]	61.68 ± 2.72	51.60 ± 1.64	49.10 ± 1.52	48.85 ± 2.27	82.04 ± 1.44	51.93 ± 1.71	13.50
	FEDformer [ICML'22]	57.20 ± 9.47	52.38 ± 6.09	49.04 ± 7.26	47.89 ± 8.44	82.13 ± 4.17	52.31 ± 7.03	13.50
	Nonformer [NeurIPS'22]	70.56 ± 0.55	61.57 ± 0.66	57.75 ± 0.72	59.10 ± 0.66	88.32 ± 0.36	63.40 ± 0.79	10.25
	PatchTST [ICLR'23]	73.23 ± 0.25	65.70 ± 0.64	60.82 ± 0.76	62.61 ± 0.34	89.74 ± 0.19	67.32 ± 0.22	4.00
	Crossformer [ICLR'23]	73.30 ± 0.14	65.06 ± 0.35	61.23 ± 0.33	62.59 ± 0.14	90.02 ± 0.06	67.43 ± 0.22	4.00
	iTransformer [ICLR'24]	69.28 ± 0.22	59.59 ± 0.45	54.62 ± 0.18	56.20 ± 0.19	86.71 ± 0.10	60.27 ± 0.21	12.00
	MTST [AISTATS'24]	72.14 ± 0.27	63.84 ± 0.72	60.01 ± 0.81	61.43 ± 0.38	88.97 ± 0.33	65.83 ± 0.51	7.00
	Medformer [NeurIPS'24]	72.87 ± 0.23	64.14 ± 0.42	60.60 ± 0.46	62.02 ± 0.37	89.66 ± 0.13	66.39 ± 0.22	6.00
	SoftShape [ICML'25]	73.94 ± 0.20	65.67 ± 0.20	61.97 ± 0.52	62.97 ± 0.52	90.59 ± 0.09	68.46 ± 0.24	2.17
	BioFormer [ICML'26]	73.37 ± 0.39	65.62 ± 0.52	60.81 ± 0.36	62.56 ± 0.30	90.37 ± 0.12	67.61 ± 0.35	3.83
	NeighSpec	74.17 ± 0.21	66.61 ± 0.55	62.84 ± 0.69	64.21 ± 0.44	90.70 ± 0.09	69.26 ± 0.36	1.00
BCI-2a (4 classes)	Transformer [NeurIPS'17]	41.68 ± 2.28	42.86 ± 1.69	41.68 ± 2.28	41.17 ± 2.50	68.88 ± 1.34	43.37 ± 1.85	8.17
	Reformer [ICLR'20]	42.69 ± 0.91	43.05 ± 1.18	42.69 ± 0.91	42.52 ± 0.94	69.64 ± 1.20	44.15 ± 1.56	3.67
	Informer [AAAI'21]	42.52 ± 1.10	43.24 ± 1.21	42.52 ± 1.10	42.24 ± 0.99	69.20 ± 1.27	43.72 ± 1.69	5.50
	Autoformer [NeurIPS'21]	27.43 ± 0.72	27.33 ± 0.82	27.43 ± 0.72	27.05 ± 0.79	53.04 ± 0.43	27.16 ± 0.31	14.00
	FEDformer [ICML'22]	29.06 ± 1.00	29.17 ± 1.12	29.06 ± 1.00	28.72 ± 1.04	54.58 ± 0.25	28.75 ± 0.19	13.00
	Nonformer [NeurIPS'22]	42.67 ± 0.63	43.60 ± 0.95	42.67 ± 0.63	42.51 ± 0.70	69.31 ± 0.49	43.98 ± 0.67	4.17
	PatchTST [ICLR'23]	40.64 ± 1.11	40.78 ± 1.29	40.64 ± 1.11	40.13 ± 1.30	67.63 ± 0.76	41.05 ± 0.88	9.83
	Crossformer [ICLR'23]	42.48 ± 2.90	43.95 ± 3.47	42.48 ± 2.90	41.61 ± 3.05	69.22 ± 2.75	43.44 ± 3.16	5.83
	iTransformer [ICLR'24]	33.58 ± 1.44	33.88 ± 1.35	33.58 ± 1.44	33.23 ± 1.47	59.57 ± 0.75	32.87 ± 0.87	12.00
	MTST [AISTATS'24]	39.31 ± 0.99	39.72 ± 0.77	39.31 ± 0.99	39.06 ± 1.11	66.61 ± 0.55	39.79 ± 0.46	10.83
	Medformer [NeurIPS'24]	42.40 ± 0.69	42.90 ± 0.72	42.40 ± 0.69	42.29 ± 0.72	69.09 ± 0.23	43.51 ± 0.31	6.67
	SoftShape [ICML'25]	40.87 ± 7.22	41.68 ± 7.35	40.87 ± 7.22	38.69 ± 8.67	68.57 ± 6.07	44.23 ± 8.13	8.33
	BioFormer [ICML'26]	46.68 ± 1.56	47.24 ± 1.64	46.68 ± 1.56	46.35 ± 1.59	72.49 ± 1.12	47.63 ± 1.53	2.00
	NeighSpec	56.65 ± 1.49	57.37 ± 1.30	56.65 ± 1.49	56.28 ± 1.47	81.02 ± 0.64	60.63 ± 1.12	1.00
BCI-2b (2 classes)	Transformer [NeurIPS'17]	70.86 ± 1.22	70.96 ± 1.28	70.86 ± 1.22	70.83 ± 1.20	77.82 ± 1.63	77.23 ± 1.74	7.67
	Reformer [ICLR'20]	70.90 ± 0.46	71.13 ± 0.37	70.90 ± 0.46	70.82 ± 0.50	78.34 ± 0.80	77.73 ± 0.98	6.67
	Informer [AAAI'21]	70.57 ± 0.68	70.64 ± 0.66	70.57 ± 0.68	70.54 ± 0.69	77.86 ± 0.70	77.07 ± 0.73	9.33
	Autoformer [NeurIPS'21]	58.00 ± 0.47	58.04 ± 0.51	58.00 ± 0.47	57.95 ± 0.44	61.06 ± 0.62	59.46 ± 0.73	14.00
	FEDformer [ICML'22]	60.99 ± 1.64	61.18 ± 1.66	60.99 ± 1.64	60.82 ± 1.68	66.19 ± 2.48	65.37 ± 2.57	13.00
	Nonformer [NeurIPS'22]	70.71 ± 0.82	71.02 ± 0.79	70.71 ± 0.82	70.60 ± 0.86	78.79 ± 0.92	78.37 ± 0.92	7.00
	PatchTST [ICLR'23]	73.26 ± 0.65	73.61 ± 0.69	73.26 ± 0.65	73.17 ± 0.68	81.69 ± 0.33	81.36 ± 0.41	2.00
	Crossformer [ICLR'23]	70.58 ± 1.21	71.16 ± 1.57	70.58 ± 1.21	70.39 ± 1.14	77.54 ± 1.03	76.88 ± 0.76	9.00
	iTransformer [ICLR'24]	65.94 ± 0.61	66.15 ± 0.68	65.94 ± 0.61	65.84 ± 0.58	72.17 ± 1.26	71.53 ± 1.43	12.00
	MTST [AISTATS'24]	71.53 ± 0.56	71.63 ± 0.61	71.53 ± 0.56	71.50 ± 0.55	79.48 ± 0.36	79.27 ± 0.45	4.00
	Medformer [NeurIPS'24]	69.40 ± 0.35	69.66 ± 0.33	69.40 ± 0.35	69.30 ± 0.36	76.03 ± 0.45	75.30 ± 0.40	11.00
	SoftShape [ICML'25]	71.28 ± 1.02	71.61 ± 0.95	71.28 ± 1.02	71.16 ± 1.08	78.48 ± 1.66	77.88 ± 1.86	5.33
	BioFormer [ICML'26]	72.83 ± 1.12	72.92 ± 1.08	72.83 ± 1.12	72.81 ± 1.13	80.38 ± 1.04	79.98 ± 1.22	3.00
	NeighSpec	77.44 ± 0.72	77.50 ± 0.70	77.44 ± 0.72	77.43 ± 0.72	85.85 ± 0.75	85.87 ± 0.70	1.00

Table 2 presents the cross-subject classification performance of NeighSpec against 13 baselines across six biomedical time-series datasets. Overall, NeighSpec achieves the best average rank of 1.25 across 36 evaluation metrics and obtains the best dataset-level average ranking on five of six benchmarks, indicating competitive cross-subject performance on both EEG and ECG data.

Performance on sample-level classification tasks. Notably, NeighSpec achieves particularly pronounced improvements on BCI-2a and BCI-2b, where different segments from the same subject correspond to different motor imagery classes. On BCI-2a, NeighSpec improves macro-F1 from 46.35% to 56.28%, an absolute gain of 9.93 percentage points, while increasing AUROC from 72.49% to 81.02%. Similarly, on BCI-2b, macro-F1 increases from 72.81% to 77.43%, accompanied by an AUROC improvement from 80.38% to 85.85%. Unlike subject-level diagnostic tasks, where labels remain constant within an individual, motor imagery classification requires distinguishing different task states despite substantial subject-dependent variation. The improvements on these datasets suggest that neighbourhood-referenced modulation may be particularly beneficial when discriminative information is associated with local spectral relationships and oscillatory patterns. By incorporating neighbouring frequency components into the correction process, NeighSpec provides a mechanism for adapting spectral representations while limiting unnecessary changes to local structures.

Performance on subject-level classification tasks. On PTB-XL, NeighSpec improves macro-F1 over BioFormer by 1.65 percentage points. On APAVA and PTB, its mean macro-F1 is also higher, by 1.39 and 2.56 percentage points, respectively, although

these differences lie within one standard deviation across seeds; on APAVA, BioFormer obtains higher Precision and AUROC. On ADFTD, BioFormer achieves the best result on all six metrics, although the macro-F1 difference is small (0.43 percentage points). These results indicate that the benefits of neighbourhood-referenced modulation vary across datasets and task settings. Nevertheless, the consistent improvements on both motor imagery benchmarks provide encouraging evidence for the relevance of local spectral relationships in cross-subject representation learning.

5.2 Spectral Structural Analysis

To investigate whether NeighSpec enhances task-discriminative spectral representations while reducing subject-dependent variation, we employ Frequency-Band Discriminability (FBD), which measures the ratio of inter-class spectral differences to intra-class variability across subjects. A higher FBD indicates stronger task-related spectral separation relative to subject-specific variation. The detailed formulation is provided in Appendix D.

Figure 4 compares the latent spectra of Transformer, Transformer+GRL, SoftShape and NeighSpec on APAVA. Transformer+GRL denotes a Transformer trained with an additional subject classifier connected through a gradient reversal layer (GRL), which adversarially penalises subject-predictive features [19]; it is used only as a reference in this analysis and in Figure 5. Notably, NeighSpec exhibits stronger spectral discriminability in the selected low-frequency region, achieving an average FBD of 3.778, compared with 3.214 for Transformer, 3.250 for Transformer+GRL, and 1.549 for SoftShape. The enlarged view further reveals a pronounced discriminative response around a normalised frequency of 0.02, suggesting that NeighSpec captures task-relevant spectral differences more effectively in this region.

These observations are consistent with our motivation that local spectral relationships provide useful information for cross-subject representation learning. By incorporating neighbourhood-referenced correction, NeighSpec can adapt spectral components according to their surrounding frequency context. Importantly, the observed improvement is frequency-dependent rather than uniform across the entire spectrum. The results therefore provide supporting representation-level evidence, while the contribution of neighbourhood-relative modulation is further examined through the ablation studies.

5.3 Effectiveness of Neighbourhood-Referenced Spectral Modulation

To examine the effectiveness of the proposed spectral modulation, we conduct controlled ablations on APAVA, PTB, and ADFTD, as summarised in Table 3.

Neighbourhood-relative correction. We first investigate whether explicitly modelling the relative discrepancy between a frequency component and its local neighbourhood improves spectral adaptation. Replacing bounded neighbourhood-relative refinement with direct magnitude convolution (Direct Magnitude Conv.) reduces F1 by 2.17, 4.95, and 1.75 percentage points on APAVA, PTB, and ADFTD, respectively. These consistent drops suggest that the effectiveness of spectral modulation depends

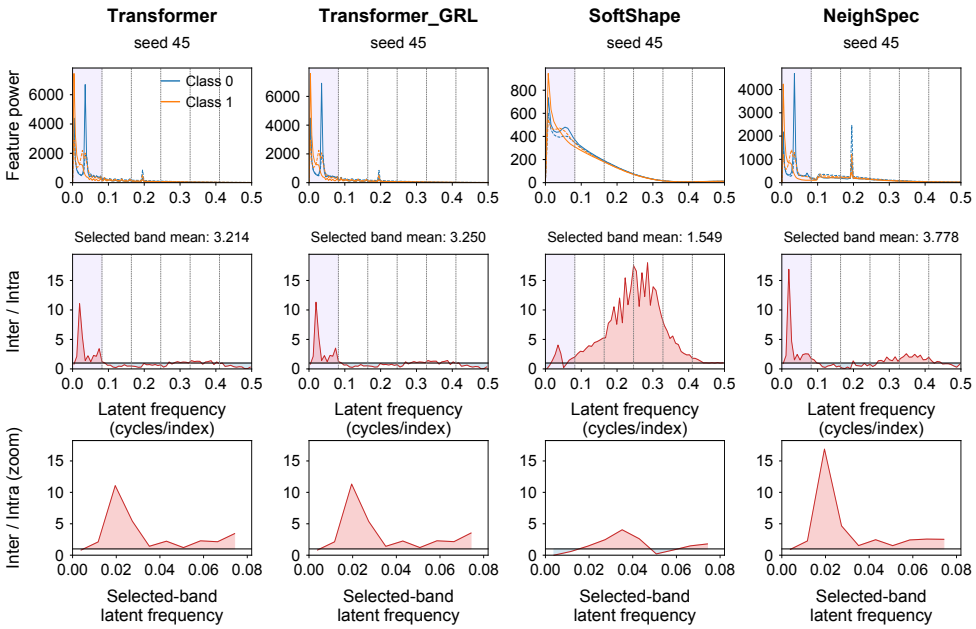

Fig. 4 Latent spectral discriminability on APAVA for Transformer, Transformer+GRL, SoftShape and NeighSpec (seed 45). From top to bottom: class-averaged latent power, FBD over the entire normalised frequency range, and a magnified low-frequency window. Horizontal positions are measured in cycles per latent index, not physical EEG hertz.

not merely on incorporating neighbouring frequencies, but also on how their relative information is used to control the correction.

Sample-conditioned modulation. We further examine whether spectral corrections should adapt to individual samples rather than rely on fixed parameters. Replacing the dynamic modulation with static band parameters (Static Band Modulation) decreases F1 by 2.44, 2.77, and 4.15 percentage points across the three datasets, respectively. These results highlight the importance of sample-dependent spectral adaptation for cross-subject generalisation.

Cross-band contextual modelling. The Content-only Attention and w/o Band Position variants investigate the role of frequency-band position information in the modulation controller. Removing positional information from the query, key, and value representations (Content-only Attention) reduces F1 across all three datasets, with the largest drop of 2.70 percentage points on PTB. Completely removing band-position embeddings (w/o Band Position) produces smaller and dataset-dependent changes, including a negligible F1 difference of 0.05 percentage points on APAVA. Although Content-only Attention achieves slightly higher Accuracy on ADFTD, the complete model obtains the highest F1 on all three datasets.

Overall, these results support the effectiveness of combining bounded neighbourhood-relative correction with sample-conditioned spectral modulation. The improvements

are most consistent for the magnitude correction mechanism and adaptive parameter generation, while the contribution of positional information varies across datasets.

Table 3 Ablation study of neighbourhood-referenced spectral alignment. Accuracy and macro-F1 (%, mean ± std) across five random seeds (41–45).

Variant	APAVA		PTB		ADFTD	
	Acc	F1	Acc	F1	Acc	F1
Direct Magnitude Conv.	81.55 ± 1.75	78.99 ± 2.54	85.85 ± 3.30	82.32 ± 4.85	53.55 ± 1.20	51.64 ± 2.18
Static Band Modulation	81.17 ± 2.77	78.72 ± 3.22	87.53 ± 1.98	84.50 ± 2.88	51.13 ± 2.37	49.24 ± 1.85
Content-only Attention	82.21 ± 1.26	80.18 ± 2.02	87.56 ± 2.28	84.57 ± 3.36	54.51 ± 1.15	52.91 ± 1.64
w/o Band Position	82.95 ± 0.70	81.11 ± 1.26	88.69 ± 1.26	86.23 ± 1.66	53.03 ± 1.28	51.91 ± 1.08
NeighSpec	83.03 ± 1.36	81.16 ± 1.96	89.54 ± 2.05	87.27 ± 2.74	54.24 ± 0.94	53.39 ± 0.89

Direct Magnitude Conv. replaces the bounded relative magnitude update with direct convolution. Content-only Attention excludes band-position information from the attention Q/K/V projections but retains it in the residual controller.

5.4 Subject Information in Learned Representations

We further investigate how subject-specific information is represented in the learned features through subject probing and clustering analysis (Table 4). On PTB, NeighSpec exhibits lower subject separability than the compared models across Subject F1, NMI, and AMI. On APAVA, the results are more mixed: NeighSpec achieves lower Subject F1 than Transformer and SoftShape, while its NMI and AMI remain higher than those of Transformer.

Notably, subject identity remains highly predictable across all evaluated representations, suggesting that cross-subject generalisation does not necessarily require the complete removal of individual information. Together with the classification improvements in Table 2, these observations are consistent with task-oriented spectral adaptation rather than strict subject-invariant learning. Since the diagnostics are obtained from independently selected checkpoints, we interpret them as descriptive evidence rather than a controlled comparison of subject invariance.

Table 4 Subject information in frozen test representations. Subject classification F1 (%), NMI, and AMI are reported for selected checkpoints. Lower values indicate weaker subject separability.

Dataset	Model	Seed	Subject F1 (%) ↓	NMI ↓	AMI ↓
APAVA	NeighSpec	45	97.95	0.4626	0.4614
APAVA	SoftShape	41	98.42	0.4827	0.4815
APAVA	Transformer	42	99.32	0.3011	0.2995
PTB	NeighSpec	44	99.24	0.5973	0.5830
PTB	SoftShape	43	99.77	0.6919	0.6810
PTB	Transformer	41	99.79	0.7712	0.7632

The t-SNE plots in Figure 5 provide a visual complement to these measurements, although their appearance depends on the visualisation procedure and selected checkpoints. The plots alone cannot establish subject invariance.

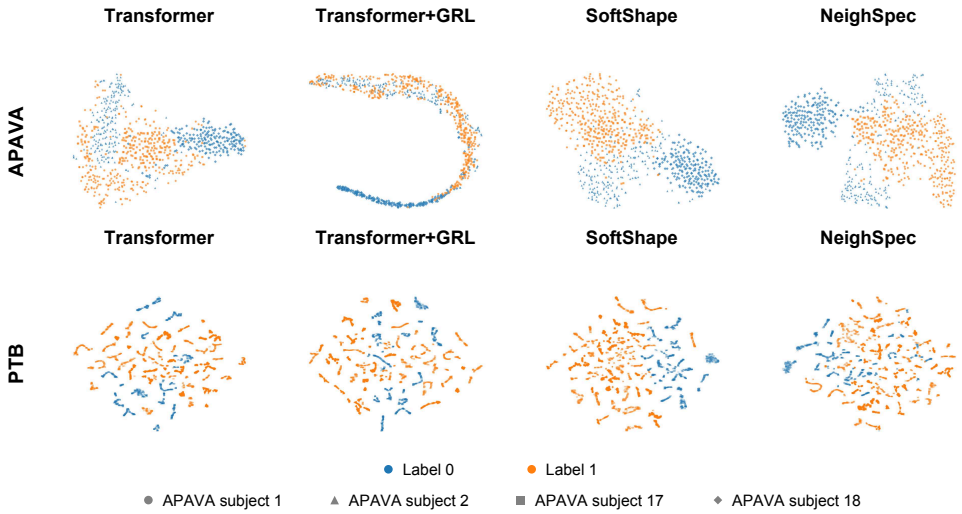

Fig. 5 t-SNE visualisation of frozen test representations on APAVA and PTB. Colours denote task labels and markers identify subjects. Transformer, Transformer+GRL, SoftShape and NeighSpec are shown using independently selected checkpoints; visual separation does not by itself establish subject invariance.

5.5 Ablation Studies

Table 5 varies magnitude modulation, phase modulation, and the ordered temporal readout. The final row corresponds to the full NeighSpec model, and the third switch controls whether the ordered temporal readout is enabled. Relative to the variant that retains the readout but disables both spectral components, the full model improves F1 from 77.95 to 81.16 on APAVA, 48.82 to 53.39 on ADFTD, and 86.20 to 87.27 on PTB.

The effects of individual components vary across datasets. Magnitude-only modulation attains 81.72 on APAVA, phase-only attains 87.49 on PTB, and joint modulation without the ordered temporal readout attains 88.16 on PTB, all above the corresponding full-model mean where applicable. The full model achieves the highest ADFTD F1 in Table 5. These results indicate interactions between spectral controls and readout rather than a monotonic benefit from every component on every dataset. This ablation focuses on the spectral modulation and temporal readout components of NeighSpec.

Table 5 Component ablations of NeighSpec. Macro-F1 (%, mean ± std) over seeds 41–45.

Mag.	Phase	Ordered Readout	APAVA	ADFTD	PTB
×	×	×	78.00 ± 1.21	47.15 ± 1.01	87.81 ± 3.42
×	×	✓	77.95 ± 4.70	48.82 ± 1.43	86.20 ± 1.42
✓	✓	×	75.18 ± 3.24	46.11 ± 1.64	88.16 ± 2.15
✓	×	✓	81.72 ± 2.94	50.90 ± 2.22	83.85 ± 2.32
×	✓	✓	80.18 ± 2.72	50.85 ± 1.21	87.49 ± 0.84
✓	✓	✓	81.16 ± 1.96	53.39 ± 0.89	87.27 ± 2.74

Table 6 Cross-subject comparison with DMMR and BioFormer. Accuracy and macro-F1 (%) are reported for four biomedical datasets. DMMR and BioFormer results are taken from [21].

Dataset	DMMR Acc.	DMMR F1	BioFormer Acc.	BioFormer F1	NeighSpec Acc.	NeighSpec F1
APAVA	64.74	63.45	82.31	79.77	83.03	81.16
PTB	77.06	68.31	87.73	84.71	89.54	87.27
ADFTD	52.21	40.83	56.73	53.82	54.24	53.39
PTB-XL	69.53	53.14	73.37	62.56	74.17	64.21

5.6 Comparison with a Domain-Generalisation Baseline for Biomedical Signals

Table 6 compares NeighSpec with DMMR [12] and BioFormer on four datasets. NeighSpec outperforms the DMMR reference results across these datasets; BioFormer is included to contextualise the spectral-alignment family. Because these measurements were not obtained under a newly unified training protocol, the comparison does not isolate differences in architecture or pretraining.

5.7 Sensitivity to Ordered Readout Bins

Figure 6 varies the number of ordered temporal readout bins over $1, 2, 4, 8, 16, 32, 64, 128$ on APAVA and reports F1 means and standard deviations over seeds 41–45. This experiment studies the temporal resolution retained by the ordered readout. Changing K changes the flattened representation dimension KD, the readout normalisation, and the following classifier's parameter count. It therefore measures a combined resolution/capacity trade-off, rather than the effect of temporal granularity at a fixed classifier size. No monotonic trend or cross-dataset optimum is inferred from this single-dataset sensitivity study.

5.8 Computational Complexity

We further evaluate the computational efficiency of NeighSpec in terms of model parameters and operation counts on APAVA, as summarised in Table 7. Compared with BioFormer, NeighSpec reduces the parameter count from 1.36M to 0.817M, a

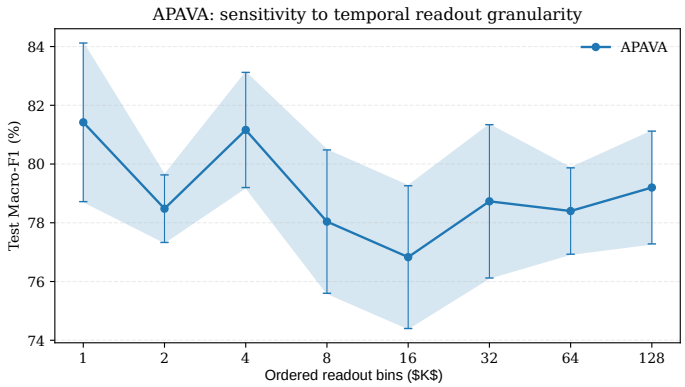

Fig. 6 Sensitivity of APAVA macro-F1 to the number of ordered readout bins ($K = 1, 2, 4, 8, 16, 32, 64, 128$). Mean $\pm$ standard deviation over seeds 41–45. Changing K also changes the classifier input dimension.

reduction of approximately 40%, while achieving a comparable or slightly higher macro-F1 (81.16% versus 79.77%). This is achieved with a compact architecture that replaces multi-scale temporal encoding with a single-scale residual embedding and lightweight spectral encoder blocks.

Rather than relying on a complex temporal backbone, NeighSpec performs adaptive spectral modulation through compact band descriptors and local neighbourhood operations. These results suggest that effective cross-subject representation learning does not necessarily require increasing model complexity, and that neighbourhood-referenced spectral adaptation can achieve competitive performance with substantially fewer parameters.

We additionally report operation estimates for reference. NeighSpec records 0.149 G operations under the THOP MAC-based convention. Since FFT-related operations and several functional layers are not included in this estimate, operation counts across models are not strictly comparable and should not be interpreted as end-to-end inference costs.

6 Discussion and Limitations

The largest gains of NeighSpec were obtained on the two motor-imagery datasets, where the class depends on the mental task rather than on the subject. This is consistent with the physiology of motor imagery: imagined hand or foot movements produce event-related desynchronisation and synchronisation of sensorimotor mu and beta rhythms, which appear as localised power changes within narrow frequency ranges whose exact position and amplitude vary across individuals [22]. A correction that refers each frequency coefficient to its immediate neighbours can adapt such subject-specific peaks without flattening the surrounding spectral structure. For the diagnostic datasets, where every recording of a subject shares the same label, the gains over

Table 7 Model complexity and cross-subject macro-F1 on APAVA.

Model	F1 (%)	Params (M)	Profiled Ops. (G)*
NeighSpec	81.16	0.817	0.149
Transformer	73.75	0.87	0.20
Reformer	75.93	0.77	0.18
Informer	69.47	1.07	0.20
Autoformer	68.06	3.61	0.91
FEDformer	73.51	1.46	0.20
Nonformer	69.74	0.94	0.20
PatchTST	55.97	0.93	0.41
Crossformer	68.93	5.23	0.21
iTransformer	72.30	0.83	0.01
MTST	64.01	2.59	0.57
Medformer	76.31	7.41	0.57
SoftShape	78.71	0.47	0.03
BioFormer	79.77	1.36	0.21

*Input size $(1, 256, 16)$, FP32. NeighSpec operations are counted as THOP multiply–accumulate operations and exclude FFT/IFFT, attention matrix multiplications, GELU and other unregistered functions. Counts for the other models use independently reported profiling scopes and are indicative only. NeighSpec (817,038 parameters) uses $L = 4$, $D = 128$, FFN width 256, kernels 11/7, $M = 5$ and $K = 4$.

BioFormer were smaller and, on APAVA and PTB, within the variability across seeds, while BioFormer remained better on ADFTD.

The subject-probing results indicate that NeighSpec does not remove subject identity from its representations; subject identity remains highly predictable for all models. The improvements therefore seem to come from emphasising task-relevant spectral structure rather than from strict subject invariance, in line with the view that cross-subject generalisation does not require all individual information to be discarded.

Several limitations should be noted. First, all datasets are evaluated with a single fixed subject-disjoint split, and some test sets contain only a few subjects (four for APAVA and two for BCI-2a and BCI-2b); the reported standard deviations reflect variability across random seeds, not across subject partitions. Second, baseline results are taken from previous publications that use the same splits and seeds, rather than being re-trained within a single unified pipeline. Third, the spectral and subject-probing analyses rely on selected checkpoints and are therefore descriptive.

NeighSpec is particularly effective when task-relevant information is reflected in consistent local spectral relationships across subjects, as commonly observed in physiological signals with structured oscillatory patterns. Its effectiveness may be reduced when such relationships are weak or insufficiently discriminative, particularly in tasks dominated by irregular temporal events, sparse transients, or non-periodic signal characteristics. In these cases, neighbourhood-referenced spectral modulation may provide limited benefits, as the relevant information cannot be adequately characterised

by local frequency relationships alone. Future work should therefore assess NeighSpec with leave-subject-out cross-validation and on additional recording protocols.

7 Conclusion

We presented NeighSpec, a compact method for cross-subject EEG and ECG classification that explicitly incorporates local frequency relationships into spectral alignment. Its neighbourhood-referenced module forms a local magnitude reference, measures the relative discrepancy between each coefficient and its surroundings, and combines bounded local refinement with independently controlled band gain and phase rotation. A residual temporal embedding and an ordered temporal readout connect this frequency-domain operation to task prediction while retaining coarse temporal structure.

Across six benchmarks, NeighSpec achieves an overall average rank of 1.25 and a higher mean macro-F1 than the published BioFormer results on five datasets, with particularly marked gains on BCI-2a and BCI-2b. The ablation results support sample-conditioned and neighbourhood-relative modulation, although the evidence does not establish that the method preserves every discriminative spectral feature or eliminates subject identity. Future studies should evaluate local spectral distortion directly, examine a broader range of recording protocols, and test how latent spectral corrections relate to physiological EEG rhythms.

Declarations

Funding. This research was supported in part by the National Natural Science Foundation of China (Grant 62372254), the Beijing–Tianjin–Hebei Basic Research Cooperation Project of Hebei Natural Science Foundation (F2024203115; 24JCZXJC00160 in Tianjin), the Innovative Development Joint Fund Key Projects of Shandong Natural Science Foundation (ZR2023LZH003 and ZR2022LZH009), and the robotic AI-Scientist platform of the Chinese Academy of Sciences.

Competing interests. The authors declare no competing interests.

Ethics approval and consent to participate. This work involves secondary analysis of previously released EEG and ECG research datasets and does not report new participant recruitment or data collection by the authors. Requirements applicable to the original recordings are governed by the respective dataset providers.

Consent for publication. Not applicable.

Data availability. All datasets used in this study are publicly available. We follow the datasets and preprocessing protocols adopted in Medformer [24] and BioFormer [21]. The datasets can be accessed through their respective public repositories.

Code availability. Code is available at https://github.com/NKU-EmbeddedSystem/NeighSpec.

Author contributions. Z.J. conducted model implementation, experiments, analysis, and the initial draft; H.L. contributed to study conception, methodology, supervision, and manuscript revision; G.D. contributed to methodological development, analysis, and revision; J.S. assisted with data preparation, evaluation, and visualisation; X.W., X.G., and J.Z. contributed to project supervision and critical review; J.S.-C. contributed

to interpretation, scientific discussion, and manuscript editing. All authors reviewed the manuscript.

References

[1] Jovic A, Frid N, Brkic K, Cifrek M. Interpretability and accuracy of machine learning algorithms for biomedical time series analysis – a scoping review. Biomedical Signal Processing and Control. 2025;110:108153.

[2] Cascarano A, Mur-Petit J, Hernández-González J, et al. Machine and Deep Learning for Longitudinal Biomedical Data: A Review of Methods and Applications. Artificial Intelligence Review. 2023;56(Suppl 2):1711–1771.

[3] Liu Z, Alavi A, Li M, Zhang X. Self-Supervised Contrastive Learning for Medical Time Series: A Systematic Review. Sensors. 2023;23(9):4221.

[4] Mushtaq F, Welke D, Gallagher A, et al. One Hundred Years of EEG for Brain and Behaviour Research. Nature Human Behaviour. 2024;8:1437–1443.

[5] Hannun AY, Rajpurkar P, Haghpanahi M, et al. Cardiologist-Level Arrhythmia Detection and Classification in Ambulatory Electrocardiograms Using a Deep Neural Network. Nature Medicine. 2019;25:65–69.

[6] Ye W, Zhang Z, Teng F, Zhang M, Wang J, Ni D, et al. Semi-Supervised Dual-Stream Self-Attentive Adversarial Graph Contrastive Learning for Cross-Subject EEG-Based Emotion Recognition. IEEE Transactions on Affective Computing. 2025;16(1):290–305.

[7] Poterucha TJ, Jing L, Ricart RP, et al. Detecting Structural Heart Disease from Electrocardiograms Using AI. Nature. 2025;644:221–230.

[8] Kopčanová M, Tait L, Donoghue T, et al. Resting-State EEG Signatures of Alzheimer's Disease Are Driven by Periodic but Not Aperiodic Changes. Neurobiology of Disease. 2024;190:106380.

[9] Chetty CA, Bhardwaj H, Kumar GP, et al. EEG Biomarkers in Alzheimer's and Prodromal Alzheimer's: A Comprehensive Analysis of Spectral and Connectivity Features. Alzheimer's Research & Therapy. 2024;16(1):236.

[10] Rayatdoost S, Yin Y, Rudrauf D, Soleymani M. Subject-Invariant Eeg Representation Learning For Emotion Recognition. In: ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing; 2021. p. 3955–3959.

[11] Zhou K, Yang Y, Qiao Y, Xiang T. Domain Generalization with MixStyle. In: International Conference on Learning Representations; 2021. .

[12] Wang Y, Zhang B, Tang Y. DMMR: Cross-Subject Domain Generalization for EEG-Based Emotion Recognition via Denoising Mixed Mutual Reconstruction. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38; 2024. p. 628–636.

[13] Roques A, Gruffaz S, Kim K, Durmus AO, Oudre L. Personalized Convolutional Dictionary Learning of Physiological Time Series. In: The 28th International Conference on Artificial Intelligence and Statistics. vol. 258; 2025. p. 1837–1845.

[14] Yin C, Yu Q, Fang Z, Peng C, Li P. Rethinking Cross-Subject Data Splitting for Brain-to-Text Decoding. In: Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing; 2025. p. 5675–5689.

[15] Kim H, Kim J, Kim SB. Cross-subject generalizable representation learning with class-subject dual labels for biosignals. Knowledge-Based Systems. 2024;295:111855.

[16] Shen X, Liu X, Hu X, Zhang D, Song S. Contrastive Learning of Subject-Invariant EEG Representations for Cross-Subject Emotion Recognition. IEEE Transactions on Affective Computing. 2023;14(3):2496–2511.

[17] Apicella A, Arpaia P, D'Errico G, Marocco D, Mastrati G, Moccaldi N, et al. Toward cross-subject and cross-session generalization in EEG-based emotion recognition: Systematic review, taxonomy, and methods. Neurocomputing. 2024;604:128354.

[18] Wen Q, Zhou T, Zhang C, Chen W, Ma Z, Yan J, et al. Transformers in Time Series: A Survey. In: Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23; 2023. p. 6778–6786.

[19] Ganin Y, Ustinova E, Ajakan H, Germain P, Larochelle H, Laviolette F, et al. Domain-Adversarial Training of Neural Networks. Journal of Machine Learning Research. 2016;17(59):1–35.

[20] Zhou Y, Hu H, Zhou Q, Guan Q, Jiang M. Rethinking Domain Generalization from Perspective of Gradient Granularity. In: Frontiers in Artificial Intelligence and Applications. vol. 392 of ECAI 2024; 2024. p. 1543–1550.

[21] Du G, Li H, Liu X, Zhang Z, Gong X, Zhang J. BioFormer: Rethinking Cross-Subject Generalization via Spectral Structural Alignment in Biomedical Time-Series. In: Proceedings of the 43rd International Conference on Machine Learning. vol. 306 of Proceedings of Machine Learning Research. PMLR; 2026. p. 26501–26528. Available from: https://proceedings.mlr.press/v306/du26f.html.

[22] Pfurtscheller G, Neuper C. Motor imagery activates primary sensorimotor area in humans. Neuroscience Letters. 1997;239(2–3):65–68. https://doi.org/10.1016/S0304-3940(97)00889-6.

[23] Nie Y, H Nguyen N, Sinthong P, Kalagnanam J. A Time Series is Worth 64 Words: Long-term Forecasting with Transformers. In: International Conference on Learning Representations; 2023. .

[24] Wang Y, Huang N, Li T, Yan Y, Zhang X. Medformer: A multi-granularity patching transformer for medical time-series classification. Advances in Neural Information Processing Systems. 2024;37:36314–36341.

[25] Liu Z, Luo Y, Li B, Eldele E, Wu M, Ma Q. Learning soft sparse shapes for efficient time-series classification. In: International Conference on Machine Learning; 2025. .

[26] Sun B, Saenko K. Deep CORAL: Correlation Alignment for Deep Domain Adaptation. In: European Conference on Computer Vision Workshops; 2016. p. 443–450.

[27] Li X, Dai Y, Ge Y, Liu J, Shan Y, DUAN L. Uncertainty Modeling for Out-of-Distribution Generalization. In: International Conference on Learning Representations; 2022. .

[28] Huang H, Si X, Han Y, Ming D. A Novel Conditional Adversarial Domain Adaptation Network for EEG Cross-Subject Emotion Recognition. IEEE Transactions on Affective Computing. 2025;16:2905–2917.

[29] Pascucci D, Menétrey MQ, Passarotto E, Luo J, Paramento M, Rubega M. EEG brain waves and alpha rhythms: Past, current and future direction. Neuroscience and Biobehavioral Reviews. 2025;176:106288.

[30] Donoghue T, Haller M, Peterson EJ, et al. Parameterizing Neural Power Spectra into Periodic and Aperiodic Components. Nature Neuroscience. 2020;23:1655–1665.

[31] Zhou T, Ma Z, Wen Q, Wang X, Sun L, Jin R. FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting. In: Proc. 39th International Conference on Machine Learning. vol. 162 of Proceedings of Machine Learning Research; 2022. p. 27268–27286.

[32] Zhou W, Zeng F, Fan T. SSMGNN: Spectral temporal graph neural network with state space models for multivariate time-series forecasting. Neurocomputing. 2026;666:132295.

[33] Escudero J, Abásolo D, Hornero R, Espino P, López M. Analysis of Electroencephalograms in Alzheimer's Disease Patients with Multiscale Entropy. Physiological Measurement. 2006;27(11):1091–1106.

[34] Miltiadous A, Tzimourta KD, Afrantou T, Ioannidis P, Grigoriadis N, Tsalikakis DG, et al. A Dataset of Scalp EEG Recordings of Alzheimer's Disease, Frontotemporal Dementia and Healthy Subjects from Routine EEG. Data. 2023;8(6).

[35] Bousseljot R, Kreiseler D, Schnabel A. Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet. Biomedizinische Technik. 1995;40:317. Supplement 1.

[36] Wagner P, Strodthoff N, Bousseljot RD, Kreiseler D, Lunze FI, Samek W, et al. PTB-XL: A Large Publicly Available ECG Dataset. Scientific Data. 2020;7:154.

[37] Brunner C, Leeb R, Müller-Putz GR, Schlögl A, Pfurtscheller G.: BCI Competition 2008 – Graz Data Set A. Graz University of Technology, dataset description. Available from: https://www.bbci.de/competition/iv/desc_2a.pdf.

[38] Leeb R, Brunner C, Müller-Putz GR, Schlögl A, Pfurtscheller G.: BCI Competition 2008 – Graz Data Set B. Graz University of Technology, dataset description. Available from: https://www.bbci.de/competition/iv/desc_2b.pdf.

[39] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention Is All You Need. In: Advances in Neural Information Processing Systems; 2017. .

[40] Kitaev N, Kaiser L, Levskaya A. Reformer: The Efficient Transformer. In: International Conference on Learning Representations; 2020. .

[41] Zhou H, Zhang S, Peng J, Zhang S, Li J, Xiong H, et al. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 35; 2021. p. 11106–11115.

[42] Wu H, Xu J, Wang J, Long M. Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting. In: Advances in Neural Information Processing Systems; 2021. .

[43] Liu Y, Wu H, Wang J, Long M. Non-stationary Transformers: Exploring the Stationarity in Time Series Forecasting. In: Advances in Neural Information Processing Systems; 2022. .

[44] Zhang Y, Yan J. Crossformer: Transformer Utilizing Cross-Dimension Dependency for Multivariate Time Series Forecasting. In: International Conference on Learning Representations; 2023. .

[45] Liu Y, Hu T, Zhang H, Wu H, Wang S, Ma L, et al. iTransformer: Inverted Transformers Are Effective for Time Series Forecasting. In: International Conference on Learning Representations; 2024. .

[46] Zhang Y, Ma L, Pal S, Zhang Y, Coates M. Multi-resolution Time-Series Transformer for Long-term Forecasting. In: Proceedings of The 27th International Conference on Artificial Intelligence and Statistics. vol. 238; 2024. p. 4222–4230.

[47] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research

resource for complex physiologic signals. Circulation. 2000;101(23):e215–e220.

[48] Tangermann M, Müller KR, Aertsen A, Birbaumer N, Braun C, Brunner C, et al. Review of the BCI Competition IV. Frontiers in Neuroscience. 2012;6:55. https://doi.org/10.3389/fnins.2012.00055.

[49] Long M, Cao Y, Wang J, Jordan MI. Learning Transferable Features with Deep Adaptation Networks. In: International Conference on Machine Learning; 2015. p. 97–105.

[50] Tan W, Zhang H, Wang Y, Wen W, Chen L, Li H, et al. SEDA-EEG: A semi-supervised emotion recognition network with domain adaptation for cross-subject EEG analysis. Neurocomput. 2025;622:129315. https://doi.org/10.1016/j.neucom.2024.129315.

Appendix Overview

Appendix A describes the NRSA implementation and forward computation, Appendix B details the experimental protocol, Appendix C compares architectural and domain-generalisation strategies, and Appendix D defines the frequency-band discriminability metric.

A Module Design and Forward Computation

A.1 Residual Temporal Embedding

NeighSpec preserves the input sequence length. After one randomly selected training augmentation, a linear channel projection maps C input channels to D latent channels. The temporal residual applies channel-wise layer normalisation, a same-length $D \to D$ convolution, GELU and dropout, and a second same-length convolution; an outer dropout precedes residual addition. Both temporal kernels are positive odd integers with symmetric zero padding. Equation (1) gives the exact order.

A.2 Neighbourhood-Referenced Spectral Alignment

For each sample and block, an orthonormal real FFT is computed along time. Only non-DC and non-Nyquist positive-frequency coefficients enter the M nonempty bands. The five descriptor entries are log1p mean amplitude, log1p channel-averaged population standard deviation, log1p channel-averaged peak amplitude, log1p mean power, and normalised energy centroid, exactly as in Eq. (2). A descriptor projection and band-position vector form each band token; single-head self-attention supplies context across bands. The context predicts D log-gains, phase angles, and refinement rates per band, plus a length-three softmax kernel shared over channels. All controls are computed separately for each sample. The magnitude reference is computed with replicate-padded sliding windows and weighted sums inside each band.

Algorithm 1 Forward Pass of Neighbourhood-Referenced Spectral Alignment (NRSA)

Require: Temporal features $H \in \mathbb{R}^{B\times T\times D}$; number of frequency bands M; modulation bounds G, ρ, Φ.
Ensure: Spectrally aligned features $H' \in \mathbb{R}^{B\times T\times D}$.

I. Spectral Decomposition.
$Z \leftarrow \mathrm{rFFT}(H)$.
Preserve the DC and Nyquist components; partition the remaining spectrum into M bands $\{B_m\}_{m=1}^{M}$.
II. Band-wise Spectral Representation.
for $m = 1, \ldots, M$ **do**
  Compute band magnitude $A_m = |Z_m|$.
  Extract spectral statistics:
  $\mathbf{s}_m \leftarrow [\log(1+\mu_m), \log(1+\sigma_m), \log(1+a_m^{\max}), \log(1+\mathcal{E}_m), c_m]$.
  $\mathbf{e}_m \leftarrow \mathrm{LN}(\mathrm{GELU}(\mathrm{Linear}(\mathbf{s}_m))) + \mathbf{p}_m$.
end for
III. Cross-Band Interaction.
$\mathbf{O} \leftarrow \mathrm{CrossBandAttn}(\mathbf{e}_1, \ldots, \mathbf{e}_M)$.
IV. Adaptive Modulation Parameters.
for $m = 1, \ldots, M$ **do**
  $w_m \leftarrow \mathrm{Softmax}(f_w(\mathbf{o}_m))$.
  $g_m \leftarrow G\tanh(f_g(\mathbf{o}_m))$.
  $r_m \leftarrow \rho\tanh(f_r(\mathbf{o}_m))$.
  $\beta_m \leftarrow \Phi\tanh(f_\beta(\mathbf{o}_m))$.
end for
V. Neighbourhood-Referenced Spectral Modulation.
for $m = 1, \ldots, M$ **do**
  Construct the local spectral reference:
  $\widetilde{A}_m \leftarrow \mathrm{Conv}(A_m, w_m)$.
  Apply bounded neighbourhood-relative magnitude correction:
  $A'_m \leftarrow A_m e^{g_m}\left[1 + r_m \tanh\left(\frac{\widetilde{A}_m - A_m}{\max(A_m, \varepsilon)}\right)\right]$.
  Apply band-wise phase rotation:
  $P'_m \leftarrow P_m \odot e^{\mathrm{i}\beta_m}$, where $P_m = e^{\mathrm{i}\arg(Z_m)}$.
  $Z'_m \leftarrow A'_m \odot P'_m$.
end for
VI. Spectral Reconstruction.
Reassemble corrected bands with preserved DC and Nyquist components.
$H' \leftarrow \mathrm{irFFT}(Z'; n = T)$.
return H'.

The encoder inserts $\mathcal{D}(H) - H$ into its dropout residual branch, followed by layer normalisation and a pointwise residual feed-forward network (Eq. (6)). Zero initialisation of the gain, phase, and refinement heads makes $\mathcal{D}$ the identity at initialisation. It does not make the complete encoder block the identity, because normalisation and the feed-forward network still operate. Similarly, DC/Nyquist preservation applies to $\mathcal{D}$ alone. The ordered readout projects features with a linear layer and GELU, applies adaptive average pooling into K ordered time bins, flattens the bins, and applies layer normalisation before the classification MLP.

B Experimental Setup

B.1 Dataset Preprocessing

All evaluations use disjoint subject partitions: every segment from an individual is assigned exclusively to training, validation, or testing. We consider clinical subject-level diagnoses (APAVA, ADFTD, PTB, and PTB-XL) together with sample-level motor-imagery classes (BCI Competition IV 2a and 2b). The statistics below describe the processed experimental subsets, not necessarily the full public archives.

APAVA [33] contains resting-state EEG from 23 participants (12 diagnosed with Alzheimer's disease and 11 controls). For the experiments, sixteen channels sampled at 256 Hz are organised into five-second trials. Standardised trials are divided into one-second windows with 50% overlap, yielding 5,967 segments. The fixed evaluation subjects are {1,2,17,18}; subjects {15,16,19,20} are used for validation, and all remaining subjects for training.

ADFTD [34] includes routine EEG of 88 participants with Alzheimer's disease, frontotemporal dementia, or no reported dementia. Nineteen-channel recordings sampled originally at 500 Hz are band-pass filtered at 0.5–45 Hz and resampled to 256 Hz. Each recording is divided into non-overlapping one-second segments, providing 69,752 examples after preprocessing. Subject-disjoint training, validation, and test groups use a 60:20:20 allocation.

PTB [35, 47] is derived from the Physikalisch-Technische Bundesanstalt ECG database [47]. We use the 198-participant myocardial-infarction/healthy-control subset represented by 15 ECG leads. Recordings originally sampled at 1,000 Hz are resampled to 250 Hz and normalised. Individual beats are extracted around detected R peaks, with outlier rejection and zero-padding to a common length; the processed subset contains 64,356 beats. Subjects are assigned to 60:20:20 train/validation/test partitions.

PTB-XL [36] contains 12-lead diagnostic ECG records. The processed cohort excludes cases with conflicting diagnostic labels across recordings, leaving 17,596 individuals. Signals from the 500-Hz release are resampled to 250 Hz and divided into one-second non-overlapping segments, yielding 191,400 samples for five-class classification. We use subject-level 60:20:20 partitions.

BCI Competition IV 2a [37, 48] comprises four motor-imagery classes for nine participants, recorded with 22 monopolar EEG electrodes at 250 Hz. For the present experiments, the C3, Cz, and C4 electrode signals are retained and the data are downsampled fourfold. Each participant completed two sessions with 288 trials per session. The processed set contains 5,184 examples; subjects {6,8} serve as validation participants and {7,9} as test participants. The remaining participants form the training set.

BCI Competition IV 2b [38, 48] records two-class left- versus right-hand imagery for nine participants using three bipolar EEG derivations around C3, Cz, and C4. The dataset includes two screening sessions without online feedback and three subsequent sessions with feedback. The processed 6,520 segments follow the same subject split as BCI-2a, with participants {6,8} for validation and {7,9} for testing.

B.2 Baseline Models

The comparative methods cover conventional sequence attention, forecasting-inspired Transformer variants, medical time-series architectures, and shape-based classification. Each is evaluated or cited in the cross-subject comparison as documented in Table 2.

Transformer [39] uses self-attention to combine sequence-wide dependencies. **Reformer** [40] replaces dense attention with locality-sensitive hashing to reduce its cost. **Informer** [41] uses sparse attention and compressed intermediate sequences for long-horizon modelling. **Autoformer** [42] combines decomposition

with autocorrelation-based sequence interactions, while **FEDformer** [31] integrates seasonal/trend decomposition and Fourier-domain processing.

Non-stationary Transformer [43] applies series stationarisation together with destationary attention to recover informative temporal dependencies. **Crossformer** [44] employs dimension-segment-wise embedding and a two-stage attention scheme for cross-time and cross-dimension interactions. **PatchTST** [23] feeds time patches to a Transformer encoder; **iTransformer** [45] instead treats variables as tokens. **MTST** [46] combines resolutions for sequence modelling, and **Medformer** [24] introduces multi-granularity patch representations for biomedical time series. **SoftShape** [25] uses learned sparse shape patterns for classification.

BioFormer [21] provides the closest spectral-alignment reference. Its published six-dataset results are shown as a separate comparator; its multi-scale architecture and normalisation are distinct from those implemented by NeighSpec.

B.3 Training and Implementation Details

NeighSpec is trained end-to-end with the cross-entropy loss. Subject identities are used only to construct the subject-disjoint splits and are never provided to the model. All experiments were implemented in PyTorch and run on NVIDIA RTX 3090 GPUs. We use the Adam optimiser with a learning rate of 10^{-4} and train for at most 100 epochs. Training stops early when the validation macro-F1 has not improved for ten consecutive epochs, and the checkpoint with the highest validation macro-F1 is evaluated on the test subjects. The batch size is 32 for APAVA, BCI-2a, and BCI-2b, and 128 for ADFTD, PTB, and PTB-XL. Each configuration is trained with five random seeds (41–45), and we report the mean and standard deviation over these runs.

Architecture. Across all six datasets, the spectral encoder uses a latent width of $D = 128$, a feed-forward width of 256, and a dropout rate of 0.1. The number of encoder layers L, the two temporal convolution kernel lengths, the number of frequency bands M, and the number of ordered readout bins K vary by dataset. In this order, the settings are $(4, (11, 7), 5, 4)$ for APAVA, $(6, (3, 3), 4, 16)$ for ADFTD, $(6, (11, 7), 3, 8)$ for PTB, $(1, (15, 11), 5, 4)$ for PTB-XL, $(4, (5, 3), 6, 8)$ for BCI-2a, and $(2, (11, 7), 3, 8)$ for BCI-2b.

Data augmentation. During training, one augmentation is drawn at random from a dataset-specific set for each batch and applied to all samples in that batch: {`scale`(0.2), `mask`(0.05)} for APAVA, {`scale`(0.05), `mask`(0.1)} for ADFTD, {`frequency`(0.1)} for PTB, {`jitter`(0.05)} for PTB-XL, {`channel`(0.25)} for BCI-2a, and {`frequency`(0.05)} for BCI-2b. Here, `jitter`(σ) adds Gaussian noise with standard deviation σ, while `scale`(σ) multiplies the signal by a random factor drawn around 1 with standard deviation σ. No augmentation is applied during validation or testing.

C Comparison with Domain Generalisation Methods

Adversarial alignment, statistical matching, and feature perturbation provide different ways to address domain shift [11, 19, 26, 27, 49]. BTS-specific methods include DMMR [12], SEDA-EEG [50], and CADAN [28]. NeighSpec instead applies a sample-conditioned, bounded spectral transform based on local frequency relationships. Table 8

compares representative architectures in terms of multi-scale modelling and explicit neighbourhood-relative correction. These architectural attributes characterise the methods; they are not performance measurements.

Table 8 Qualitative comparison of representative time-series models in terms of architectural characteristics and spectral neighbourhood modelling.

Model	Multi-scale	Multi-granularity	Frequency-aware	Designed for BTS	Spectral Neighbourhood	Relative Correction
Transformer [39] / Reformer [40] / Informer [41]	×	×	×	×	×	×
Autoformer [42] / FEDformer [31]	✓	×	✓	×	×	×
Crossformer [44] / PatchTST [23] / MTST [46]	✓	✓	×	×	×	×
iTransformer [45]	×	×	×	×	×	×
Medformer [24]	✓	✓	×	✓	×	×
BioFormer [21]	✓	✓	✓	✓	✓	×
NeighSpec	×	×	✓	✓	✓	✓

Spectral Neighbourhood indicates explicit interactions between adjacent frequency coefficients. Relative Correction denotes modulation guided by the normalised discrepancy between a frequency component and its local spectral reference. The comparison focuses on explicitly designed mechanisms rather than implicitly learned dependencies.

D Frequency-Band Discriminability (FBD)

FBD is used only as a post hoc descriptive statistic of class-related spectral separation, not as a training loss. It compares the spread of mean spectral power between classes against its variation between participants within each class. The method follows the diagnostic introduced in BioFormer [21], with the representation coordinates interpreted in the latent rather than calibrated physical-frequency domain.

For a temporal embedding $z(t) \in \mathbb{R}^D$, define the power spectrum of latent channel d by $\mathrm{PSD}_d(f) = |\mathcal{F}\{z_d(t)\}(f)|^2$. Let $\mathcal{B}$ be a partition of the positive-frequency coordinates into equal-width intervals, indexed by b. The average power of an interval is

$$E_b = \frac{1}{D} \sum_{d=1}^{D} \int_{f \in b} \mathrm{PSD}_d(f)\, df. \tag{8}$$

Averaging over segments yields $\overline{E}_b^{(y,s)}$ for class y and subject s. Let N_{cls} denote the number of classes and S_y the number of subjects in class y. Define the class-level mean and the across-subject within-class dispersion as

$$\mu_b^{(y)} = \frac{1}{S_y} \sum_{s} \overline{E}_b^{(y,s)}, \qquad \mathrm{Intra}(b) = \frac{1}{N_{\mathrm{cls}}} \sum_{y=1}^{N_{\mathrm{cls}}} \mathrm{Std}_s\left(\overline{E}_b^{(y,s)}\right). \tag{9}$$

The range of class means measures the corresponding between-class discrepancy:

$$\mathrm{Inter}(b) = \max_y \mu_b^{(y)} - \min_y \mu_b^{(y)}. \tag{10}$$

FBD is then defined by

$$\mathrm{FBD}(b) = \frac{\mathrm{Inter}(b)}{\mathrm{Intra}(b) + \varepsilon}, \quad \varepsilon > 0. \tag{11}$$

A larger value indicates stronger measured class separation relative to the measured subject dispersion in that interval. Near-zero within-class variation requires care: the stabiliser prevents a division by zero but cannot rule out an unstable large ratio. Summaries over wider regions average these interval-level ratios. For the matched-seed APAVA visualisation, the horizontal coordinates are normalised latent frequencies in $[0, 0.5]$ cycles per index; they are not physical EEG frequencies.